specialized diagnostic arrays significantly improves the accuracy of diagnosis of suspect masses in the pancreas. We have constructed diagnostic arrays to only contain genes with diagnostic and/or prognostic potential for the classification of pancreatic tissues, augmented with control features. Our results demonstrate that this setup is suitable to produce reliable, reproducible and informative expression profiles of pancreatic tissues and biopsy samples. Expression profiling analysis using the specialized diagnostic array in conjunction with conventional cytology allows the distinction between pancreatic ductal adenocarcinoma (PDAC) and non-malignant diseases of the pancreas with almost 100% diagnostic accuracy. We are currently in the process of analyzing additional tumor entities, such as acinar and neuroendocrine tumors, using both the diagnostic array as well as large scale arrays, in order to develop a multiclass classification system for the comprehensive diagnosis of different malignancies in the pancreas. In addition, we expect further development of the array in combination with careful analysis of clinical patient data to result in the recognition of distinct prognostic gene expression signatures predicting important clinical parameters such as stage of disease, response to therapy, or prognosis. Specialized DNA arrays thus represent valuable new diagnostic tools which can significantly expand the range of information gained in routine diagnostic procedures, thus providing a better basis for decisions on treatment options and setting the stage for therapeutic regimens custom tailored to the individual patient.

doi:10.1016/j.ejcsup.2006.04.016

S16. HIERARCHICAL NEURAL NET TECHNOLOGY FOR MOLECULAR STAGING WITH CLINICAL DATA

R. Kates[a,b], H. Allgayer[c], N. Harbeck[b]. [a]*REK Consulting, Germany;* [b]*Technical University of Munich, Germany;* [c]*Department of Experimental Surgery, Universitätsklinikum Mannheim, Ruprecht-Karls-University Heidelberg, Mannheim, Germany.*

Trained neural networks can be used to construct scoring models for molecular staging in cancer. They offer considerable flexibility for representing nonlinear interactions, but because of their flexibility they also tend to require substantial "training" data. It often happens that clinical factors are available in a large collective, but molecular data is available only for a smaller subset. A hierarchical neural net training architecture is presented here. Hierarchical nets are trained in levels, the first level on a large cohort with limited (usually clinical) factors, the second and possibly higher levels on cohort subsets with more (usually molecular) factors, and so on. The scores produced at the first level are treated as "factors" for the second level, and so on. In diseases with distinctly classifiable modes or sites of recurrence (e.g., bone vs. soft tissue in breast cancer), the "competing" risks can be modelled within the neural network architecture. To test the hypothesis that a molecular staging factor might signal the particular relapse mode, one can study significant correlations between factors and "hidden" nodes of a trained neural network. A method is also described for using trained neural nets to generate hypotheses about potential subgroups for molecular staging targets. Applications to breast, colon, and gastric cancer are reviewed.

doi:10.1016/j.ejcsup.2006.04.017

S17. BIOINFORMATICS TOOLS FOR MOLECULAR CANCER DIAGNOSTICS ON MICROARRAYS

Jörg Hoheisel. *Functional Genome Analysis, German Cancer Research Center, Heidelberg, Germany.*

In the *Division of Functional Genome Analysis* at DKFZ, we are developing technologies for the identification, description and evaluation of cellular functions and their regulation by producing and processing biological information on a genomic scale. Many chemical and biophysical issues are being addressed in an attempt to understand the underlying procedural aspects, thereby establishing superior analysis processes.

Concerning human material, systems are being developed toward early diagnosis, prognosis and evaluation of the success of disease treatment with an accentuation on cancer. To this end, comparative studies on epigenetic and splice variations, transcription factor binding, transcriptional activity and actual protein expression are under way. Early diagnosis from body fluids is being worked at that is based on the binding of their components to peptide and antibody microarrays.

Combining this data with clinical information permits the definition of patient sub-groups and may provide a robust means for diagnosis and prognosis and lead to the identification of relevant molecular activities. We have established processes – both experimentally and in the area of bioinformatics – to deal with this challenge. The combination will not only occur in silico, since the various molecular levels affect each other extensively. Soon, current in silico systems biology will translate into ever more complex experimental set-ups that permit an evaluation of a biological issue in a systemic experimentation. Similar to research in physics, an iterative interaction of theoretical and experimental systems biology will yield important insights into function, providing the archetypal platform for an eventual model of a cell.

doi:10.1016/j.ejcsup.2006.04.018

S18. CONVENTIONAL STATISTICAL METHODOLOGY IN LARGE SCALE PROFILING

Rainer Spang. *Max Planck Institute for Molecular Genetics, Berlin, Germany.*

A plethora of computational methods for the diagnosis of cancer using gene expression profiling has been suggested. This might come as a surprise, since diagnosis appears to be a straightforward classification problem. What can be done with methods from standard statistics textbooks? Clearly, the problem is the large number of genes on the arrays. Including them all in a classification model leads to saturation of the model.